Probing molecular free energy landscapes by periodic loading

Oliver Braun,[1] Andreas Hanke,[1, 2] and Udo Seifert[1]

[1]*II. Institut für Theoretische Physik, Universität Stuttgart, 70550 Stuttgart, Germany*
[2]*Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK*

Single molecule pulling experiments provide information about interactions in biomolecules that cannot be obtained by any other method. However, the reconstruction of the molecule's free energy profile from the experimental data is still a challenge, in particular for the unstable barrier regions. We propose a new method for obtaining the full profile by introducing a periodic ramp and using Jarzynski's identity for obtaining equilibrium quantities from non-equilibrium data. Our simulated experiments show that this method delivers significant more accurate data than previous methods, under the constraint of equal experimental effort.

PACS numbers: 87.15.La, 87.15.Aa, 05.70.Ln, 87.64.Dz

Introduction - A key feature of biological systems is the high degree of self-organization of polymers, proteins, and other macromolecules, and their interaction with smaller components such as energy providers or messenger molecules [1]. These processes are ultimately driven by specific and tunable molecular interactions. Their detailed knowledge is thus a prerequisite for the understanding how biological systems work on molecular and higher levels. Recent developments of highly sensitive force probes such as atomic force microscopy (AFM) [2, 3], optical and magnetic tweezers [4, 5, 6], and biomembrane force probes [7, 8] make it possible to probe the molecular interactions of individual biomolecules by their response to mechanical stress (see [9, 10] for reviews). The systems studied by single molecule pulling experiments can be divided in two groups: in *rupture* experiments, receptor and ligand molecules are attached to a substrate and a transducer, respectively, often via chemical linkers. After allowing receptors and ligands to bind, the transducer, e.g., an AFM cantilever, is pulled away, which causes the receptor-ligand pairs to rupture. The maximum force the molecule can withstand has been measured in this way for biotin and streptavidin [2, 8], and many other receptor-ligand pairs [11]. Secondly, *un- and refolding* experiments probe the elastic properties of an individual biomolecule. The molecule is attached between a substrate and a transducer, again via chemical linkers. Force-extension relations are obtained by measuring the force as a function of the position of the transducer. In this way one may explore regions of the free energy landscape of the biomolecule far away from thermal folding pathways. Investigated systems include DNA [12, 13, 14], RNA [15, 16], polysaccharides [17], the muscle protein titin [18], the membrane protein bacteriorhodopsin (BR) [19], and many other proteins [20].

Figure 1 shows a typical setup of single molecule pulling experiments. The molecule is attached between the substrate surface and the cantilever tip. The position of the cantilever $x(t)$ is moved according to a prescribed experimental protocol. The extension of the molecule is described by a suitable reaction coordinate z given by the position of the cantilever tip. For fixed extension z
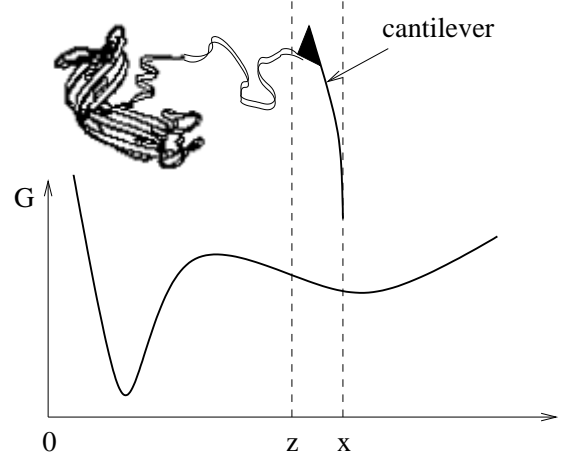

Figure 1: Schematic view of the experimental setup and a generic free energy potential $G(z)$. The first minimum represents the folded state, whereas the second shallow minimum represents the unfolded state of the biopolymer. The coordinate x denotes the position of the cantilever and z the position of the cantilever tip to which one end of the biopolymer is attached.

and time t, the energy of the molecule perturbed by the cantilever spring is given by the time-dependent Hamiltonian

$$H(z,t) = G(z) + V_0(z, x(t)) \equiv G(z) + \frac{k}{2} (x(t) - z)^2 ,$$
(1)

where $G(z)$ is the free energy profile of the unperturbed molecule. The second term describes the external force acting on the molecule in terms of a harmonic potential with effective spring constant k. Since the molecule is coupled to a heat bath at temperature T the time evolution of z is stochastic.

Traditionally, the cantilever is moved according to a linear ramp,

$$x(t) = x_0 + vt ,$$
(2)

with offset x_0 at $t = 0$ and constant velocity v, and the force $F(t, v)$ acting on the cantilever is recorded. The

challenge is to recover from these data the unperturbed molecule's free energy profile $G(z)$ containing the desired information about the molecular interactions. Evans and Ritchie first pointed out that the rupture force of receptor-ligand pairs depends on the loading rate v [21]. Thus, by combining the data $F(t,v)$ for a broad spectrum of loading rates v, referred to as dynamic force spectroscopy, important features of $G(z)$ can be determined such as the distance between the minimum and maximum of an energy barrier for rupture [21]. Heymann and Grubmüller refined this technique and obtained the heights and positions of the maxima of a molecular force profile $\partial_z G(z)$ with high spatial resolution [22]. On the other hand, traditional experimental protocols like the linear ramp (2) still entail certain drawbacks. First of all, the thermodynamically unstable (concave) barrier regions of $G(z)$, determined by specific molecular interactions and therefore of particular interest, are poorly sampled due to snapping motion and thus hard to determine [22].

Periodic loading - In an effort to improve the quality of data obtained by single molecule pulling experiments, in this work we propose a new method for obtaining the full free energy profile $G(z)$ by introducing a *periodic* ramp,

$$x(t) = x_0 + a\sin(\omega t) \ , \qquad (3)$$

with given offset x_0, amplitude a, and frequency ω. Figure 2 shows that periodic loading delivers significant more accurate data for the sample free energy profile $G(z)$ than the linear ramp (2), under the constraint of equal experimental effort. The improvement of the quality of data in the important barrier region of $G(z)$ is striking. The better performance of the periodic loading method as compared to linear loading is mainly due to the fact that periodic loading ensures that the barrier region is traversed often and from both sides. The quality of our reconstruction moreover depends crucially on the fact that we sample the barrier region under non-equilibrium conditions taking advantage of Jarzynski's relation to recover the equilibrium profile [23, 24]. Driving the system out of equilibrium is important since under quasi-static conditions an efficient sampling of the barrier region (where $G(z) \gg k_B T$) is inhibited by the equilibrium Boltzmann factor $\exp[-G(z)/(k_B T)] \ll 1$. For periodic loading, the freedom to choose the frequency ω large enough allows one to probe the same region under non-equilibrium conditions, thus overriding the exponential punishment by the equilibrium Boltzmann factor. The optimal frequency arises from balancing competing effects as quantified in a case study below. This frequency should not be too large in order to enable the system to follow the external drive. Moreover, the Jarzynski procedure converges the slower the further one is away from equilibrium [25]. For too small ω, on the other hand, one does not generate enough crossings under the constraint of a finite total measuring time.
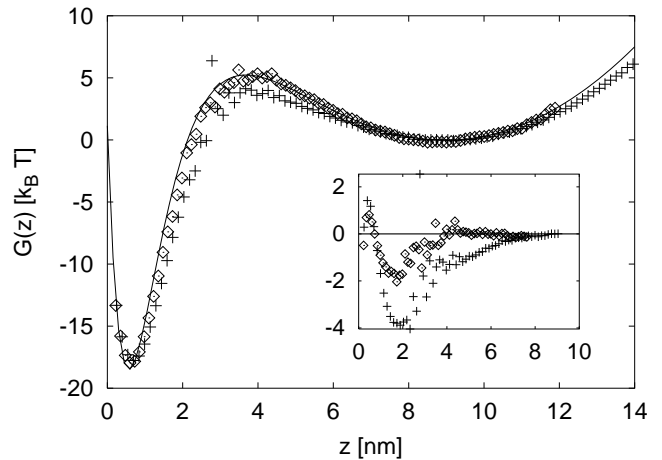

Figure 2: Comparison of reconstructed free energy profiles $G(z)$ by using periodic ($\Diamond$) vs linear loading ($+$), generated by (6). The solid line is the original free energy profile $G(z)$. For both methods, 10 trajectories of 75 ms length, a spring constant $k = 11.6$ pN/nm, and a diffusion constant $D = 10^{-7} \mathrm{cm}^2/\mathrm{s}$ were used [see (1) and (4)]. For periodic loading we used the optimal frequency $\omega^* = 1.2 \times 10^3$ 1/s (determined in Fig. 3), the preloading offset $x_0 = 6$ nm, and the amplitude $a = 5$ nm in (3). For linear loading we used the optimal velocity $v^* = 200$ nm/s (determined in Fig. 3). The inset compares the relative deviation of the reconstructed data from the original profile $G(z)$ corresponding to the zero line.

Simulated experiments - We have tested our proposal of periodic loading with simulated experiments and compared it with the traditional method of linear loading. To this end we have chosen a generic free energy profile $G(z)$ for the unfolding of tertiary structures of biopolymers such as the membrane protein BR [19]. Our sample free energy profile has two separated minima, one of which is narrow and deep representing the folded state and one of which is shallow representing the unfolded state, see Fig. 1. For BR, a rich structure of unfolding transitions under force was found [19]. Single force peaks in the unfolding spectra could be allocated to specific changes in molecular configuration. Most of the force peaks scatter between 25 up to 100 pN. With a typical length scale of several nanometers this yields an energy barrier of about 20 $k_B T$ at room temperature. We focus on one of such transitions and choose a barrier of 2 nm length and 23 $k_B T$ height, leading to a typical transition force of about 50 pN in our simulated pulling experiments.

For comparing the periodic with the linear ramp, we simulated both kinds of protocols using typical parameters as given in the caption of Fig. 2. We have generated an ensemble of trajectories $\zeta(t)$ of the reaction coordinate z by discretizing the Langevin equation

$$dz/dt = -\gamma^{-1}\partial H(z,t)/\partial z + \sqrt{2D}\,\xi(t) \qquad (4)$$

with the Hamiltonian $H(z,t)$ from (1). The Gaussian random force $\xi(t)$ has zero mean and short-ranged tem-

poral correlations $\langle \xi(t)\xi(t')\rangle = \delta(t-t')$. The diffusion constant D is related to the friction coefficient γ by the Einstein relation $D = k_B T/\gamma$. In our simulations, the length of a time step is limited by the condition that a spatial step should be small compared to the typical length scale set by the free energy profile. In addition, the recording rate of the data $\zeta(t)$ should be much smaller than an inverse time step but large enough to resolve the cantilever motion.

The reconstructed free energy profile for both protocols is shown in Fig. 2. The overall quality of the data obtained by the periodic ramp is far better than the linear ramp, especially in the barrier region where the data obtained by the linear ramp underestimates the barrier height by several $k_B T$. In order to ensure an unbiased comparison, we have chosen the same total number of trajectories and the same total measuring time for both methods. In practice, the measuring time of the trajectories, once prepared, is not a limiting factor. The periodic ramp therefore allows measuring as many transitions as necessary to collect the sufficient amount of data.

Reconstruction using Jarzynski's relation - Pulling protocols both in real experiments and in simulations used here typically generate non-equilibrium data from which one has to recover an equilibrium property like $G(z)$. Furthermore, as outlined above, the method proposed here purposely takes advantage of the non-equilibrium conditions generated by a large enough optimal driving frequency ω^*, to be determined below. The difficulty to recover equilibrium properties from non-equilibrium

data may be resolved by using a recent advance in non-equilibrium statistical mechanics due to Jarzynski [23], according to which the equilibrium profile $G(z)$ can be inferred by suitably averaging non-equilibrium trajectories $\zeta(t)$ of the reaction coordinate z [24]. This method was verified by stretching RNA reversibly and irreversibly between two conformations indeed [26].

Jarzynski's relation, in general, states that the free energy difference ΔG between two equilibrium states can be extracted from averaging the work W required to drive the system from one state to the other according to $e^{-\beta \Delta G} = \langle e^{-\beta W}\rangle$ with $\beta = 1/k_B T$ the inverse temperature [23]. The generalization from two states to a z-resolved free energy profile $G(z)$ perturbed by a harmonic spring (1) reads [24]

$$e^{-\beta[H(z,t)-G_0]} = \left\langle \delta[z-\zeta(t)]\, e^{-\beta W(t)}\right\rangle . \qquad (5)$$

The average $\langle ...\rangle$ is over infinitely many realizations $\zeta(t')$, $0 < t' < t$, of the stochastic trajectory of the biopolymer's end position, starting in equilibrium at $t' = 0$ and ending at the given position z at $t' = t$, as enforced by the delta function. The external work is a functional of $\zeta(t')$ and given by $W(t) = \int_0^t dt'\, \partial_\tau H(\zeta(t'),\tau)|_{\tau=t'}$. The constant $G_0 = -k_B T \ln(\int_0^\infty dz'\, e^{-\beta H(z',0)}/\lambda_T)$ is the free energy in the initial state at $t = 0$, where the thermal wavelength $\lambda_T = h/\sqrt{2\pi m k_B T}$ serves for normalization.

Summing up the normalized distributions obtained from (5) at each time slice with the method of weighted histograms [27] yields the reconstruction formula for the unperturbed free energy profile of the molecule

$$G(z) = -\beta^{-1}\ln\sum_t \frac{\langle\delta(z-\zeta(t))\exp(-\beta W(t))\rangle}{\langle\exp(-\beta W(t))\rangle}\Big/\sum_t \frac{\exp[-\beta V_0(z,x(t))]}{\langle\exp(-\beta W(t))\rangle} . \qquad (6)$$

Using this expression, we have generated the data shown in Fig. 2.

Optimization with respect to frequency/velocity - The quality of the reconstructed free energy profile depends crucially on the driving frequency ω for the periodic ramp and the velocity v for the linear ramp, respectively. To quantify this observation, we calculate the mean square error $\sigma^2 \equiv \langle[\widetilde{G}(z) - G(z)]^2\rangle_{z_i}$ where the average is taken over discrete values z_i in the z-interval under consideration. For clarity, we denote by $\widetilde{G}(z)$ the reconstructed free energy profile based on (6). For a better reconstruction quality, σ^2 is smaller. The z-interval is chosen from the first to the second minimum. The total measuring time and the number of trajectories are the same as before.

Figure 3 shows the mean square error σ^2 of the reconstructed free energy and some characteristic confidence intervals (error bars) for both protocols. For periodic loading, the best results were obtained for the optimal driving frequency $\omega^* \simeq 1.2 \times 10^3$ 1/s, which yields an error of $\sigma^2 \simeq 0.9\,(k_B T)^2$. This frequency is somewhat smaller than the spontaneous transition rate under preloading, which is about 7×10^3 1/s for our model system. By analyzing the work distribution we have convinced ourselves that ω^* indeed corresponds to non-equilibrium conditions. For both smaller and larger frequencies than ω^* the quality of the reconstructed data becomes worse as expected from our reasoning above.

For linear loading, the error increases for increasing driving velocity v, as expected. Since we fixed the total measuring time and the total number of trajectories, the least possible velocity for overcoming the barrier is $v \simeq 100$ nm/s. The smallest error of $\sigma^2 \simeq 1.7\,(k_B T)^2$, however, was observed at a larger, optimal velocity $v^* \simeq 200$ nm/s, with an error bar of 0.03. For the unbiased
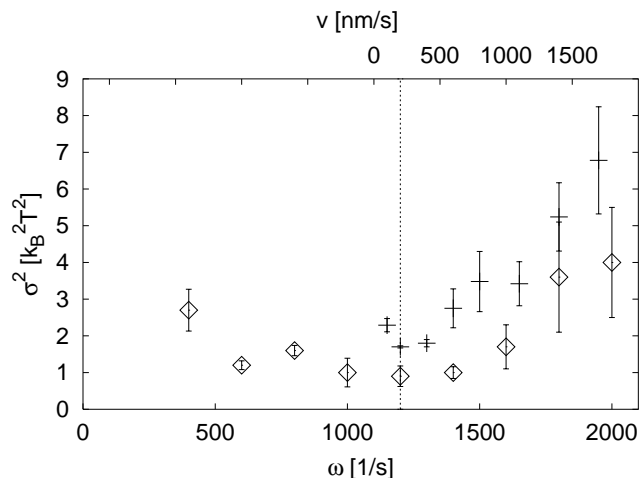
Figure 3: Mean square error σ^2 and error bars for reconstructed free energy profiles by using the linear ramp (2) as a function of v (top scale, $+$), and the periodic ramp (3) as a function of ω (bottom scale, $\Diamond$) (compare Fig. 2). Both methods use 10 runs with 10 trajectories each. The optimal velocity v^* and frequency ω^*, respectively, indicated by the vertical dashed line, correspond to the smallest error.

comparison in Fig. 2, we have chosen the optimal values ω^* and v^* from the data shown in Fig. 3.

Discussion and summary - We have proposed a new method for recovering the free energy profile $G(z)$ of biomolecules in single molecule pulling experiments by combining the new periodic ramp (3) with Jarzynski's identity for recovering equilibrium properties from non-equilibrium data. The simulated data in Fig. 2 show that the periodic ramp delivers significantly more accurate data than the traditional linear loading (2), under the constraint of equal experimental effort. An additional advantage of the periodic ramp is the fact that the measuring time may be chosen as long as necessary to collect the sufficient amount of data to recover the barrier regions of the free energy profile, which are hard to determine by previous methods.

The driving frequency ω and preloading offset x_0 of the periodic ramp (3) provide handles to optimize its performance (see Fig. 3). As our model case study has shown, the frequency should be large enough to drive the system out of equilibrium, but still smaller than the spontaneous transition rate under preloading. Whether the optimal frequency is always of this order remains to be seen in more comprehensive systematic investigations left for future work.

Our method can be extended to probing free energy profiles with different transitions between a number of metastable states. If these states are sufficiently separated, by several nanometers and energy barriers of several $k_B T$, each transition can be selected by using suitable values for preloading offset x_0 and amplitude a in (3). Subsequently, these parts of the free energy profile may be fitted together by adjusting free additive constants on each part, similarly as shown in [28].

---

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of the cell* (Garland, New York, 2002).

[2] E. L. Florin, V. T. Moy, and H. E. Gaub, Science **264**, 415 (1994).

[3] A. L. Chen and V. T. Moy, Method. Cell Biol. **68**, 301 (2002).

[4] A. D. Mehta, M. Rief, J. A. Spudich, D. A. Smith, and R. M. Simmons, Science **283**, 1689 (1999).

[5] T. R. Strick, V. Croquette, and D. Bensimon, Nature **404**, 901 (2000).

[6] C. Danilowicz, V. W. Coljee, C. Bouzigues, D. K. Lubensky, D. R. Nelson, and M. Prentiss, Proc. Natl. Acad. Sci. USA **100**, 1694 (2003).

[7] E. Evans, K. Ritchie, and R. Merkel, Biophys. J. **68**, 2580 (1995).

[8] R. Merkel, P. Nassoy, A. Leung, K. Ritchie, and E. Evans, Nature **397**, 50 (1999).

[9] R. Merkel, Phys. Rep. **346(5)**, 344 (2001).

[10] M. Rief and H. Grubmüller, ChemPhysChem **3**, 255 (2002).

[11] J. W. Weisel, H. Shuman, and R. I. Litvinov, Curr. Op. Struc. Biol. **13**, 227 (2003).

[12] J. F. Marko and E. D. Siggia, Macromolecules **28**, 8759 (1995).

[13] T. Strunz, K. Oroszlan, R. Schäfer, and H.-J. Güntherodt, Proc. Natl. Acad. Sci. USA **96**, 11277 (1999).

[14] M. C. Williams, I. Rouzina, and V. A. Bloomfield, Acc. Chem. Res. **35**, 159 (2002).

[15] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco Jr., and C. Bustamante, Science **292**, 733 (2001).

[16] S. Harlepp, T. Marchal, J. Robert, J-F. Léger, A. Xayaphoummine, H. Isambert, and D. Chatenay, preprint physics/0309063.

[17] M. Rief, F. Oesterhelt, B. Heymann, and H. E. Gaub, Science **275**, 1295 (1997).

[18] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, Science **276**, 1109 (1997).

[19] D. J. Müller, M. Kessler, F. Oesterhelt, C. Möller, D. Oesterhelt, and H. Gaub, Biophys. J. **83**, 3578 (2002).

[20] X. Zhuang and M. Rief, Curr. Op. Struc. Biol. **13**, 88 (2003).

[21] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).

[22] B. Heymann and H. Grubmüller, Phys. Rev. Lett. **84**, 6126 (2000).

[23] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997); Phys. Rev. E **56**, 5018 (1997).

[24] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. USA **98**, 3658 (2001).

[25] D. M. Zuckerman and T. B. Woolf, Phys. Rev. Lett. **89**, 180602 (2002).

[26] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., and C. Bustamante, Science **296**, 1832 (2002).

[27] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[28] M. Ø. Jensen, S. Park, E. Tajkhorshid, and K. Schulten, Proc. Natl. Acad. Sci. USA **99**, 6731 (2002).